Winemakers Dilemma - Part 2

From conversations with Alejandro about important decisions at the Winery, understanding when to harvest is among the most critical. Currently it is approaching the 40th week of the year (towards late harvest). The white grapes are ready for harvest now but will certainly yield a lower sugar content. The alternative is to wait. However, storms are frequent occurrences this time of year and bring significant risk.

If a storm occurs (weekly precipitation of >= 0.35 inches of rain and max temperature of <= 80 degrees Fahrenheit) it will spoil a significant portion of the harvest. However, if a storm does not come the grapes will ripen to a much more ideal sugar content and even have the chance to develop **botrytis mold**. This special occurrence is often referred to as 'Nobel Rot' and allows for the fermenting of the most lucrative Trockenbeerenauslese Riesling variety.

Task 1. Your data team has complied several decades of weather data relevant to the vineyard. Alejandro says in his experience and from feedback of neighbors, a storm happens about 50% of the time. Your job is to develop a model to predict if a storm will occur next week and see if you can predict better than the expert-knowledge heuristic. Here are a few guidelines:

- Only use historical data from weeks 35 to 40 of each year to control for seasonality
- Data collected is daily so it must be aggregated
- Ensure you avoid data leakage by converting the timeseries data into tabular form (t0, t-1, etc.)
- Use your choice of ML classification model but only use the sklearn library

You must use DagsHub to track progress and host this project. Complete the following sub tasks.

- Create a DagsHub account and new project
- Review the docs
- Use this project to track and push updates during your model development process
- Use DVC to track the changes in data manipulation relevant to this project
- Complete and record at least 3 experiments, tracking the metrics of accuracy, precision, and recall
- Submit a short (less than one page) written report of your model's performance on predicting storms including a link to your public DagsHub project. Discuss the following:
    - Tools and infrastructure used (or still needed)
    - The quality of the model (testing)
    - Trade-offs with model use
    - Recommended approaches to deploying and monitoring the model for future use